CONNEXIONS™

The Interoperability Report

March 1991 Volume 5, No. 3

ConneXions— The Interoperability Report tracks current and emerging standards and technologies within the computer and communications industry.

In this issue:

ConneXions is published monthly by Interop, Inc., 480 San Antonio Road, Suite 100, Mountain View, CA 94040, USA. 415-941-3399. Fax: 415-949-1779.

ISSN 0894–5926

From the Editor

The Internet Protocol (IP) has recently been extended to support *multicasting,* that is, the transmission of IP datagrams to *groups* of hosts rather than just individual hosts. IP multicasting provides IP-layer access to the multicast capability of networks such as Ethernet and FDDI, and extends that capability beyond individual networks to support internet-wide multicasting. Prior this protocol extension, most IP traffic took the form of either *unicast* messages (one datagram to each destination host) or *broadcast* messages (every host has to "listen" to traffic which is not necessarily intended for it.) IP Multicasting offers several benefits which are outlined in an article by Steve Deering of Xerox Palo Alto Research Center (PARC).

If you've followed the popular trade press recently, you've no doubt seen mention of *Frame Relay.* We asked Ed Kozel of cisco Systems to explain exactly what Frame Relay is and what a group of vendors have done in order to accelerate the introduction of this emerging technology.

Until the OSI X.400 electronic messaging standard is implemented by all vendors and on all networks, the reality will continue to consists of a number of disperate e-mail "islands." It would be nice if these systems were able to talk to each other via electronic mail gateways. The University of Southern California's Information Sciences Institute (USC-ISI) have been experimenting with such gateways for a number of years. In this issue they describe their work on the so-called *Intermail Service* and the *Commercial Mail Relay Project.* For background reading, the authors have also included short overviews of the Internet and the Domain Name System at the end of the main article.

Also in this issue, you'll find a book review and a reminder about our *1991 Internetworking Tutorials Program.* The first set of these courses will be offered in Washington, DC later this month, so call us now at 1-800-INTEROP or 415-941-3399 for more information or to sign up.

Our April issue (to be released at the IFIP *Second International Symposium on Integrated Network Management.*) will have a couple of articles related to network management. The main feature is an overview of OSI System Management. This article is yet another installment in our *Components of OSI* series. Also included is a look at how the *Simple Network Management Protocol* (SNMP) was used to control a stereo system during INTEROP 90. Stay tuned!

IP Multicasting

by Stephen E. Deering, Xerox Palo Alto Research Center

Introduction

The Internet Protocol (IP) has recently been extended to support *multicasting,* that is, the transmission of IP datagrams to *groups* of hosts rather than just individual hosts. IP multicasting provides IP-layer access to the multicast capability of networks such as Ethernet and FDDI, and extends that capability beyond individual networks to support internet-wide multicasting.

Benefits

IP multicasting offers two benefits to internet applications:

- *Efficient multi-destination delivery:* For an application that sends the same information to more than one destination—such as interactive conferencing or electronic news dissemination—multicasting is more efficient than *unicasting* separate copies to each destination: it reduces the transmission overhead on the sender and the internet and it reduces the time taken for all destinations to receive the information.
- *Robust unknown-destination delivery:* Multicasting can be used to reach one or more destinations whose individual addresses are unknown to the sender. This property can be used, for example, for locating or advertising particular internet services, such as bootstrap service or time service; for such applications, multicasting is simpler and more reliable than alternative binding mechanisms such as directory servers or configuration files.

These benefits have previously been available on some networks through the use of IP *broadcasting* [7], which is the special case of multicasting to the set of all hosts on a single network. Unfortunately, that is not a particularly useful special case—few IP datagrams are of interest to all hosts on a network (some of which might not even be running IP!), and most potential multicast applications naturally span more than one network.

The overhead imposed by IP broadcasting on uninterested hosts has made it a service to be avoided rather than exploited. The requirement that all destinations of an IP broadcast be on the same network has limited its usefulness to only a few applications, such as routing, address resolution, and cross-network bootstrapping. (In the case of one bootstrap protocol, BOOTP [2], an *ad hoc* broadcast forwarding mechanism has been developed, in order to overcome the single-network constraint). The IP multicasting service described in this article offers a more efficient, more widely applicable alternative to IP broadcasting.

The IP Multicast Service Model

The IP multicast service model, specified in RFC 1112 [5], is the natural internet analog of the type of multicast service offered by standard LANs such as Ethernet. Similar to LAN multicast addresses, there are IP multicast addresses, and a host accomplishes a multicast simply by using an IP multicast address rather than an IP individual address in the destination field of an IP datagram. Such a datagram is transmitted to all hosts that happen to be "listening" to that particular multicast address at the time the datagram is sent; the listening hosts may reside on different networks within the internet. Since IP is a "best-efforts" datagram service, there is no guarantee that all listening hosts will successfully receive a multicast transmission, or that they will all receive different multicasts in the same order; those applications that require stronger reliability or ordering properties must achieve them through the use of higher-layer protocols, as is the case with normal unicast IP service.

IP multicast addresses are encoded as *Class D* IP addresses, as illustrated in Figure 1. Note that, unlike Class A, B, and C addresses, Class D addresses do not contain a "network" subfield, since multicast destinations may span more than one network—IP multicast address are allocated from a single, flat address space. The Class D format accommodates more than 26 million different addresses, ranging from 224.0.0.0 to 239.255.255.255 in standard Internet "dotted-decimal" notation.

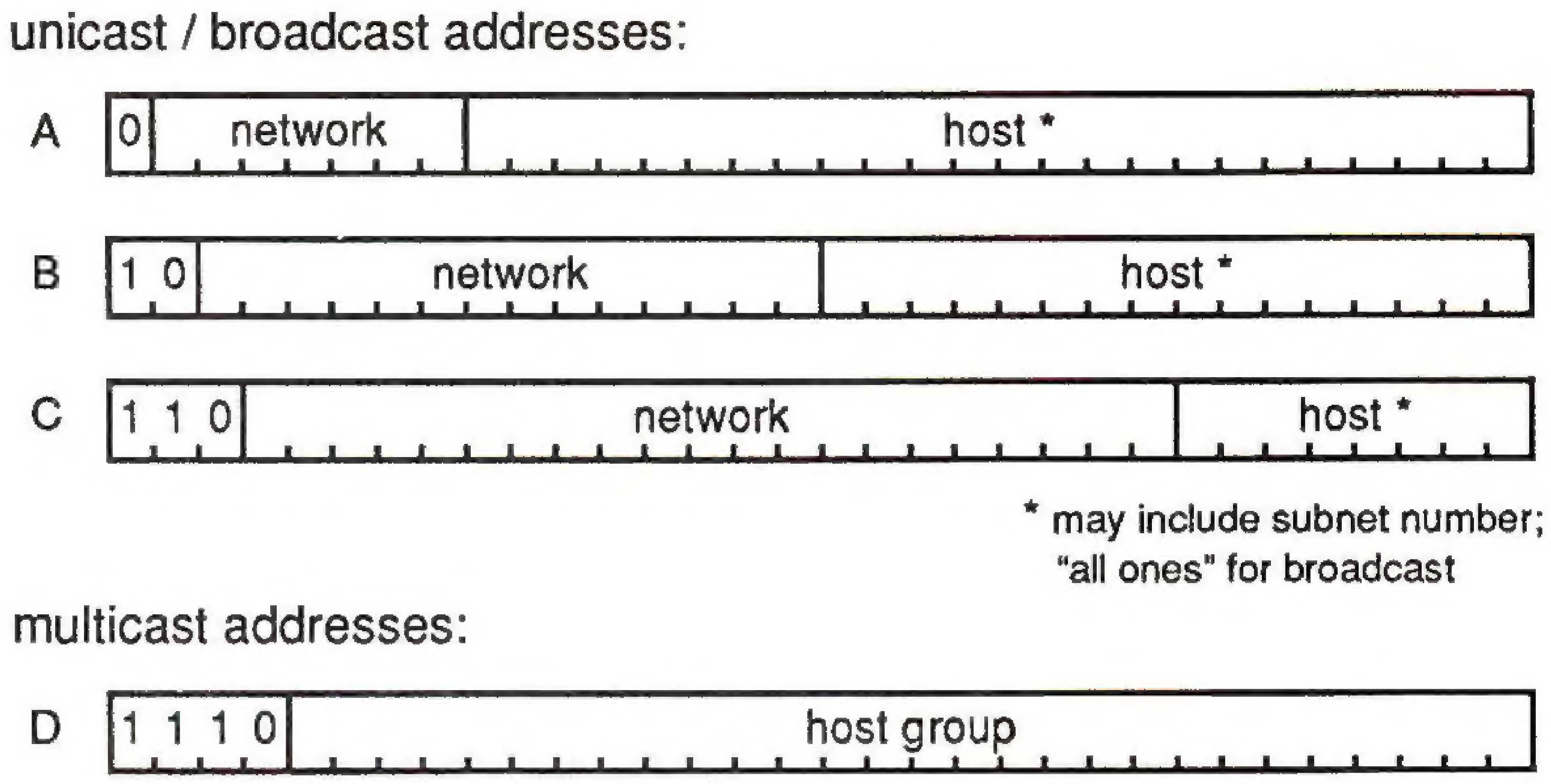


Figure 1: IP Address Formats

Host Groups

The set of hosts listening to a particular IP multicast address is called a *host group.* Hosts may join and leave host groups at will, and may belong to more than one host group at a time. There is no restriction on the number of hosts in a host group. A host need not be a member of a host group to send datagrams to it.

A host group may be *permanent* or *transient.* A permanent group has a well-known, administratively assigned IP multicast address. It is the address, not the membership, of the group that is permanent; at any time a permanent group may have any number of host members, including zero. Permanent group addresses are mainly useful as "logical addresses" for locating (or advertising) common services, where specific addresses for the service (or the clients of the service) are unknown to the sender. Since they are permanently assigned, they may be compiled into application programs or burned into ROMs, thus eliminating the need for manual configuration.

Those IP multicast addresses that are not reserved for permanent groups are available for dynamic assignment to transient groups, which are considered to exist only for as long as they have at least one member. Transient groups are useful for applications such as conferencing or distributed computing, in which a temporary association is formed among a group of hosts for the duration of a single conference or computation. After the association terminates, its group address becomes eligible for reuse by another application.

The mechanism for dynamically allocating transient group addresses is not defined or constrained by the IP multicast service model, and it is anticipated that different portions of the IP multicast address space will be allocated using different techniques. For example, there may be a number of servers that can be contacted to acquire a new transient group address. Some higher-layer protocols may generate higher-level transient "process group" or "entity group" identifiers which are then algorithmically mapped or hashed to a subset of the IP transient host group addresses.

continued on next page

IP Multicasting *(continued)*

A range of IP multicast addresses might even be set aside for random allocation by applications that can tolerate reception of unwanted datagrams from other multicast users; such applications could, perhaps, try several different multicast address from that range until a suitably "quiet" one is found.

In general, a host cannot assume that datagrams sent to any host group address will reach only the intended hosts, or that datagrams received as a member of a transient host group are intended for the recipient. Misdelivery must be detected at a layer above IP, using higher-level identifiers or authenticators. Information transmitted to an IP multicast address—like that transmitted to an IP unicast address—should be encrypted or governed by administrative routing controls if the sender is concerned about unwanted listeners.

Multicast scope control

Local-area networks, by definition, have a small geographic range, and they typically serve only a single community or administrative unit. Internetworks, on the other hand, may span the globe and serve a large number of communities and organizations. Therefore, when multicasting in an internet, unlike a LAN, it is meaningful to talk about some members of a host group being "closer" to the sender than others, either in terms of geographical or topological distance (e.g., number of network hops) or "administrative distance" (a host belonging to the same organization as the sender is administratively closer than a host belonging to a different organization). Some applications of IP multicast can benefit from the ability to limit the *scope* of a multicast transmission, in order to reach only those group members within a given distance of the sender.

There are several possible reasons for limiting the scope of a multicast:

- When using multicast to locate a particular service, such as bootstrap service or printer service, the sender may not trust, or be authorized to use, servers beyond its own administrative domain.
- Some information that is multicast may be meaningful only to nearby members of a group, for example, multicast reports of unusual network events may be of interest only to nearby members of a group of network management stations.
- When sending a query to a large group, such as an internet-wide group of directory servers, it may be preferable to reach only a few of the members, so as not to be inundated with replies and to avoid having every member service every request.

TTL

The scope of an IP multicast datagram is controlled by the *time-to-live* (TTL) field in its IP header. The TTL field limits the number of network hops that the datagram can travel. (It also, in theory, limits the number of seconds that the datagram may exist in the internet, though few routers enforce such a limit.) In order to provide meaningful administrative scope control—for example, to limit a multicast transmission to a single site—the internet routers at the boundaries of an administrative unit refuse to forward any multicasts whose TTL is less than a particular *threshold.* For example, by convention the TTL threshold for a "site" is 32; a multicast datagram sent with an initial TTL of 32 can reach any destination group members within the sender's site (assuming the site has a diameter less than 32, a conservative assumption), but cannot reach any members beyond the site.

For some applications, it is only necessary for a multicast datagram to reach *one* member of a group, for example, when using multicast to locate any one of a set of equivalent servers (e.g., directory servers). Furthermore, it is often preferable that that member be the one *nearest* the sender, in terms of delay or round-trip-time, so that subsequent interactions with that member have minimum response time. An approximation to this capability can be achieved through the use of TTL scope control, but at a finer granularity than the administrative scope boundaries discussed above. In particular, a host may perform an *expanding-ring search* for a group member by multicasting a query, starting with a TTL of one, and retransmitting with progressively larger TTLs, until a response is received.

A small range of IP multicast addresses, from 224.0.0.0 to 224.0.0.255, is reserved for those few applications that have no need to multicast further than one network hop. Such applications are those specifically concerned with routing or address management, for which individual network boundaries (as opposed to administrative boundaries) are significant. Although the number of such applications is small, the number of hosts involved in them is very large; in particular, there is a host group to which *all* hosts belong (224.0.0.1) and another for all routers (224.0.0.2). To prevent the accidental transmission of multicasts with scope wider than one hop to such groups, the internet routers refuse to forward any datagrams destined to that range of addresses, regardless of their TTLs.

Extensions to the IP service interface

An upper-layer (e.g., transport-layer) protocol sends an IP multicast datagram using the normal "Send Datagram" operation provided by the service interface to the IP layer, simply specifying an IP multicast destination address and an appropriate TTL. Similarly, IP multicasts are received via the normal "Receive Datagram" operation. However, before multicast datagrams can be received, the upper layer(s) must tell the IP layer which specific IP multicast addresses it is to listen to; this requires that the following two new operations be provided at the IP service interface:

Join Host Group (IP multicast address, interface)

Leave Host Group (IP multicast address, interface)

The Join Host Group operation requests that the host begin listening to the specified IP multicast address on a particular network interface. The Leave Host Group operation requests that the host cease listening to the identified multicast address on the given interface. The interface need not be specified in hosts that can have only one network interface; for hosts that support more than one interface, special values may be passed to indicate "all interfaces" or "a default interface." More than one upper-layer entity may ask to join the same host group, in which case, all of them must ask to leave before the IP layer will cease listening to the group's address.

How it works

Figure 2 illustrates the delivery of a multicast datagram from a source host (the circle labeled *s*) to a host group whose members (labeled *m*) are distributed across multiple networks. The source host sends the datagram as a *local network* multicast on its directly-attached network (the thick horizontal line to which *s* is connected), which causes it to be delivered to all group members attached to that network, plus any attached routers (represented by boxes).

continued on next page

IP Multicasting *(continued)*

The routers then take responsibility for delivering one copy of the datagram to every other network to which group members are attached, across an arbitrary topology of networks and routers (represented by the "cloud"). On each of those destination networks, the datagram is transmitted as a local multicast, to reach all attached group members.

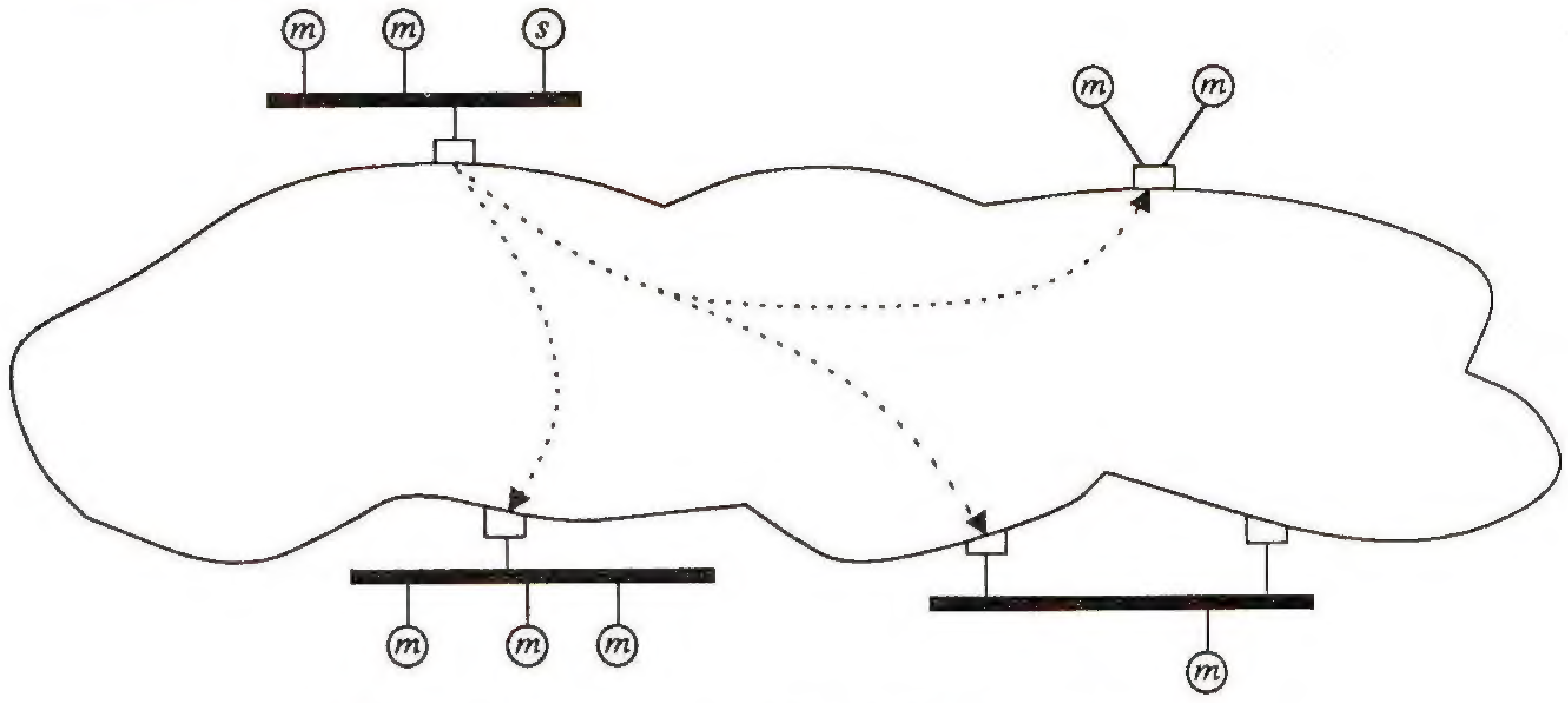


Figure 2: Internet Multicast Delivery

Address mappings

The local network multicast addresses are derived from the IP multicast address in a network-specific manner. For example, Ethernet or IEEE 802.3 multicast addresses are derived from IP multicast addresses as illustrated in Figure 3. The high-order 25 bits of the 48-bit Ethernet address are a globally unique prefix, distinguishing this use of Ethernet multicast from all other uses; the rest of the Ethernet address is taken from the low-order 23 bits of the IP multicast address. (Note that the Ethernet address is represented in so-called "canonical" bit order, in which the "group" bit is the low-order bit of the first octet.) Other networks that use the same address format as Ethernet, such as FDDI, use this same address mapping function. Networks with different local address formats require different mapping functions; for example, on a network that supports broadcast but not multicast, all IP multicast addresses may be mapped to a single local broadcast address. The specific mapping function for any particular type of network is described in the RFC that specifies how to transmit IP datagrams over that type of network (or *will* be described, in the next version of that RFC).

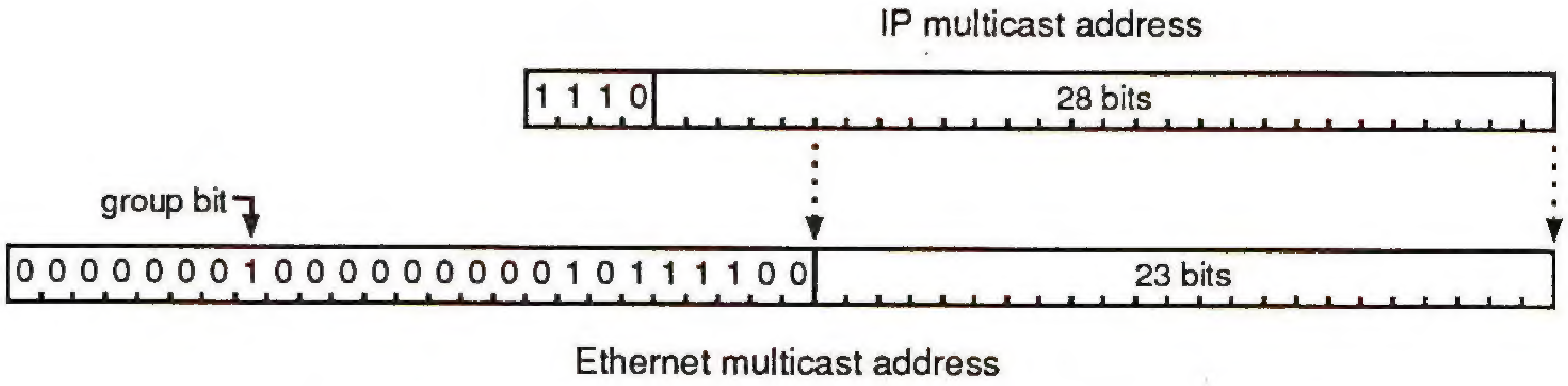


Figure 3: IP to Ethernet Multicast Address Mapping

Each host listens to those local network multicast addresses that correspond to its current IP host group memberships, as requested by the Join Host Group and Leave Host Group operations described above. Ideally, the host's network interface contains multicast address filtering hardware that can be used to prevent the reception of unwanted multicasts. In practice, many current LAN interfaces implement multicast address filtering in such a way that more than the desired addresses can slip through (e.g., all addresses that hash to the same, few values).

Worse than those are interfaces that can listen to only a small, fixed number of multicast addresses—with such interfaces, when the number of desired group memberships exceeds the filter limit, the interface must be placed in "multicast-promiscuous" mode, to receive all multicast datagrams regardless of address. Even when an interface provides perfect filtering of local multicast addresses, a host may still receive unwanted multicast datagrams on those networks where more than one IP multicast address can be mapped to the same local multicast address; this is the case with the Ethernet mapping illustrated in Figure 3, where the 28 significant bits of an IP multicast address are hashed (by truncation of high-order bits) into a 23-bit field in the Ethernet address. Thus, the IP layer must be prepared to perform software filtering of incoming multicast datagrams, based on their destination IP addresses. Imperfect hardware filtering somewhat diminishes one of the benefits of multicasting, but it's never worse than broadcasting. With the growing use of multicast (not only by IP), interface vendors can be expected to provide better multicast filters in the future.

A network that does not support either multicast or broadcast, such as a public X.25 network, can be modeled as a collection of logical point-to-point links, each joining a pair of routers or a host and a router, and each treated as a separate network for the purpose of multicast delivery. By forwarding multicasts on these logical point-to-point links, the IP routers can provide *intra*net multicasting, as well as *inter* net multicasting (albeit less efficiently than if the network itself supported multicast). The upper-right part of Figure 2 illustrates delivery of a multicast to two destination member hosts via (possibly logical) point-to-point links.

The IP routers provide loop-free forwarding of multicast datagrams from source network to destination group member networks. Multicast routing is a generalization of unicast routing, in that a datagram may have any number of destination networks, not just one, and several multicast routing algorithms have been developed as extensions of existing unicast routing algorithms. Description of those algorithms is beyond the scope of this article; the interested reader is directed to reference [6] for a full treatment of the topic.

IGMP

As with unicast routing, multicast routing is mostly a private issue between the routers, in which hosts are not involved. However, before the routers can route a datagram to a host group, they must learn where (i.e., on which networks) the members of the group reside; this information must be supplied directly by the hosts. (This problem does not arise with unicast routing, since each IP unicast address contains the number of the network on which the address resides.) Hosts report their group memberships to their neighboring routers by using the *Internet Group Management Protocol* (IGMP), which is specified in RFC 1112. Further distribution of group membership information to other routers is a private matter between the routers, as required by the specific routing algorithm in use. IGMP insulates the hosts from knowledge of the multicast algorithm, and allows different algorithms to be deployed without modifying the hosts.

Current status

The multicast extensions to IP specified in RFC 1112 have been adopted by the IAB as a *Recommended Internet Standard.* The Host Requirements RFC [1], states that hosts SHOULD support IP multicasting on all networks for which a mapping from Class D IP addresses to local multicast addresses has been defined.

continued on next page

IP Multicasting *(continued)*

At the time of writing this article, this includes Ethernet/IEEE 802.3, FDDI, SMDS, any type of point-to-point link (e.g., HDLC, SLIP, PPP), and any network that supports broadcast but not multicast. Address mappings for other types of networks will be defined in the future.

Support for the Internet Group Management Protocol (IGMP) is currently OPTIONAL, since IGMP serves no purpose on networks that have no multicast-capable routers attached, or in hosts that have no need to receive multicasts originating on other networks. It is expected that IGMP will become recommended at some future date, when multicast-capable routers have become more widely available. The Internet Assigned Numbers Authority [IANA] (Joyce Reynolds, `jkrey@isi.edu`) is accepting applications for well-known IP multicast addresses.

Availability

An implementation of IP multicasting (including IGMP) for BSD 4.3-tahoe UNIX and related systems (SunOS 4.x, Ultrix 3.x) is available by anonymous FTP from Stanford University; fetch the file `vmtp-ip/ipmulticast.README` from host `gregorio.stanford.edu` for details. This software has also been ported to several other systems, including some models of Hewlett Packard and Silicon Graphics workstations, and is expected to be included in 4.4BSD UNIX. Information about these and other implementations of IP multicasting can be obtained via the `vmtp-ip` mailing list, which you can join by sending a message to `vmtp-ip-request@gregorio.stanford.edu`.

Included in the Stanford software is an experimental multicast routing demon that implements a more recent version of the routing protocol defined in RFC 1075 [9]. This routing software may be run on any BSD-based system that has been upgraded to support IP multicasting; the system need not be a (unicast) router and need not be attached to more than one network. A “tunneling” mechanism is provided to enable multicast routing demons on different networks to forward multicast datagrams to each other through non-multicast-capable routers. This allows multicast delivery to be accomplished among any cooperating set of contiguous or non-contiguous networks within the Internet.

Stanford’s multicast routing software is based on a distance-vector routing protocol and is vulnerable to the well-known stability and scaling problems of that routing technology. It is intended only as an interim solution to the problem of providing an IP multicast routing service, to be replaced by a more robust and more scalable routing technology in the future. One effort at providing that technology is occurring in the Multicast OSPF Working Group of the Internet Engineering Task Force, which is developing a multicast extension to the OSPF routing protocol [8].

There are only a few applications that currently take advantage of IP multicast. Included in the Stanford distribution are some test programs, including multicast versions of the *ping* and *rwhod* programs. Experimental voice, video, and text conferencing applications based on IP multicast have been developed, or are being developed, at several research sites. Some researchers are investigating the use of IP multicast to improve the efficiency and configurability of existing Internet protocols, such as the *Domain Name System* (DNS), the *Network Time Protocol* (NTP) and the *Network News Transfer Protocol* (NNTP).

Any current use of IP broadcast is a candidate for migration to IP multicast, and any new protocols that would otherwise use broadcast, such as the *Router Discovery* protocol being developed by the IETF, can be expected to use IP multicast instead. And, of course, there are already several IP multicast-based games in existence, including an X11 version of the classic *Mazewar* game.

One important and active area of research is the design of multicast transport protocols, to provide reliable, sequenced multicast delivery on top of the basic IP multicast datagram service. The *Versatile Message Transaction Protocol* (VMTP) [3, 4] is one example of a such a protocol; there have been a couple of tentative proposals for a "multicast TCP." For many applications of multicast, such as resource location or voice and video conferencing, best-efforts datagram service is adequate and UDP suffices as a transport protocol.

Conclusions

IP multicasting is an important new internet service which has the potential to:

- ease the migration of existing LAN-based multicast applications and distributed systems to an internet environment,
- improve the efficiency and robustness of existing applications, and
- enable the development of entirely new classes of internet applications.

It is immediately useful as a replacement for IP broadcasting to protect hosts from receiving unwanted datagrams and, as multicast-capable IP routers are deployed, it will extend the benefits of multicast delivery beyond the confines of a single network, as IP has always done for unicast delivery.

References

[1] Braden, R., Ed., "Requirements for Internet Hosts—Communication Layers, RFC 1122.

[2] Croft, W. & Gilmore, J., "Bootstrap Protocol (BOOTP)," RFC 951.

[3] Cheriton, D., "VMTP: Versatile Message Transaction Protocol," RFC 1045.

[4] Mason, A., "VMTP: A High Performance Transport Protocol," *ConneXions,* Volume 4, No. 6, June 1990.

[5] Deering, S., "Host Extensions for IP Multicasting," RFC 1112.

[6] Deering, S. & Cheriton, D., "Multicast Routing in Datagram Internetworks and Extended LANs," ACM *Transactions on Computer Systems,* Volume 8, No. 2, May 1990.

[7] Mogul, J., "Broadcasting Internet Datagrams," RFC 919.

[8] Moy, J., "The OSPF Specification," RFC 1131.

[9] Waitzman, D., Partridge, C. & Deering, S., "Distance Vector Multicast Routing Protocol," RFC 1075.

STEPHEN E. DEERING received his B.Sc. (1973) and M.Sc. (1982) from the University of British Columbia, and expects to receive his Ph.D. from Stanford University very soon. He has been studying, designing and implementing computer communication protocols since 1978, including work on X.25 and X.400, and on high-performance transport protocols for distributed systems. For his doctoral dissertation, he has developed several new routing protocols for efficient and scalable internet multicasting. Mr. Deering has recently joined the research staff at Xerox PARC, where he is continuing his work on multicast routing and applications, and investigating new communication architectures and services. He is an active member of the Internet End-to-End Research Group and of the Internet Engineering Task Force.

The cisco/DEC/NTI/Stratacom Frame Relay Specification

by Edward R. Kozel, cisco Systems

Introduction

In an effort to speed the introduction of Frame Relay switching technology for LAN inter-connection, several switch and router vendors have proposed enhancements to the draft ANSI T1S1 Frame Relay specification. These enhancements were designed to better support LAN interconnection via Frame Relay subnetworks, and to resolve some of the problems currently experienced with LAN interconnection via X.25 [6] networks. A joint specification for enhanced Frame Relay services was recently published by the four organizations, with the dual intent of getting support for a common definition of Frame Relay services by other vendors, and proposing the enhanced services to the ANSI committee responsible for completing the definition of Frame Relay. Since publication, over seventeen companies have announced support for the Extended Frame Relay specification.

The purpose of this article is to briefly summarize the key concepts of Frame Relay, discuss the LAN/WAN interconnection issues faced when considering Frame Relay as the WAN technology, and introduce the key extensions defined in the cisco/DEC/NTI/Stratacom specification. The complete specification can be obtained at no cost from any of the four parties whose contact information is listed at the end of this article. It should be noted that the proposed extensions are initially implemented to varying degrees by the various vendors, and this article may refer at times to features only supported by some of them.

What is Frame Relay?

Frame Relay was initially defined as an additional packet service on ISDN networks. The basic architecture is presented in Recommendation I.122 of the CCITT [1], which presents the basic premise of efficient relaying of HDLC data frames. The HDLC frames are further defined in format by CCITT Recommendation Q.921/I.441 [2]. These specifications offer a service that is connection oriented and somewhat similar in nature to X.25, but without the error correction, retransmission, and recovery offered by X.25. Furthermore, service is defined at data rates far higher than X.25 networks (up to 2Mbits/second currently).

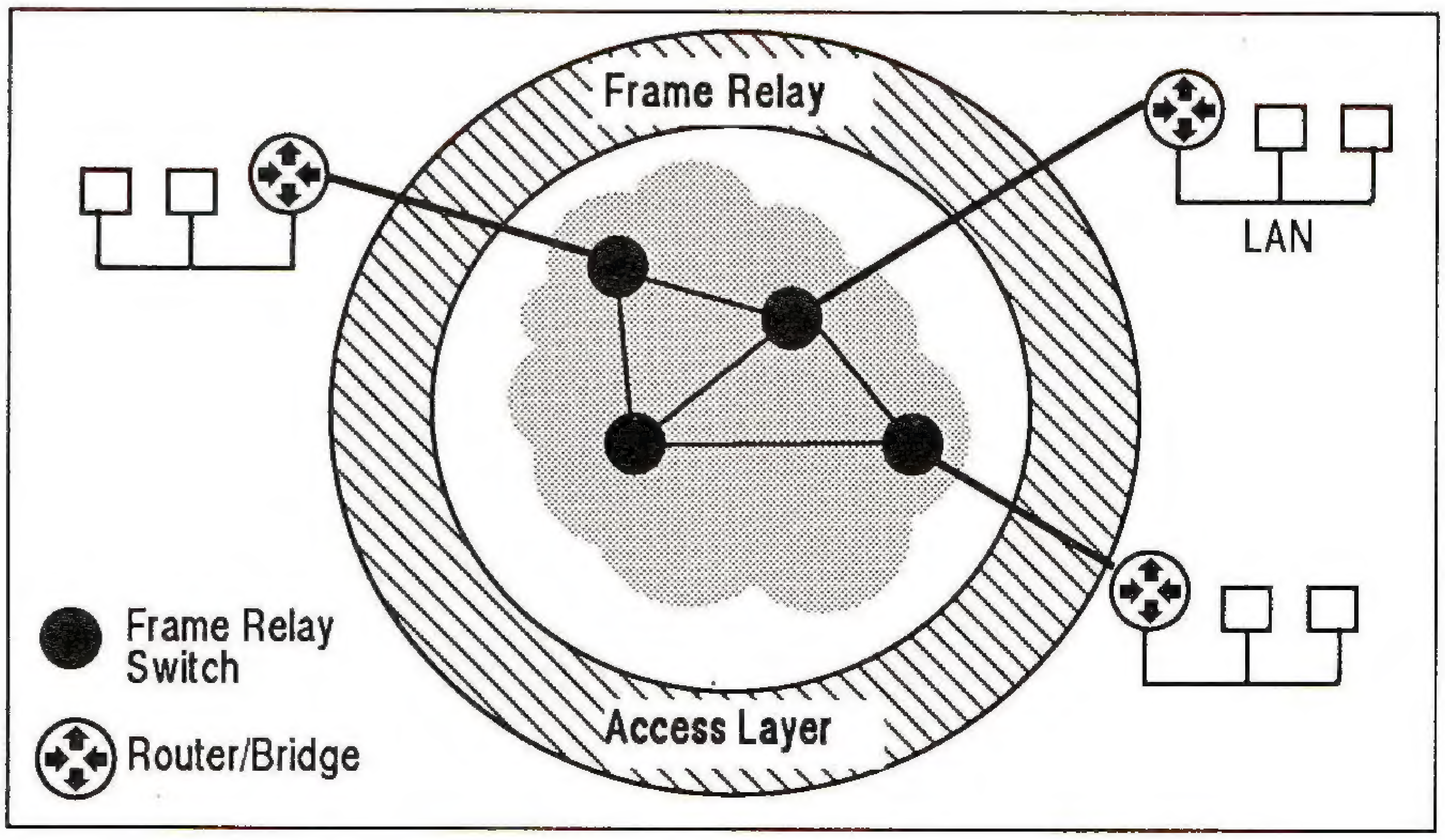


Figure 1: LAN interconnection via a Frame Relay Network

Originally envisioned as an ISDN service, Frame Relay has found much broader acceptance as a more efficient datagram-based subnetwork composed of specialized packet switches facilitating LAN interconnection. Intermediate switches in the network will forward each packet based on the address in the packet; the actual underlying protocols between switches vary widely from cell relay (Stratacom) to actual relaying of complete HDLC frames (AT&T IACS). Organizations currently utilizing X.25 networks are looking towards Frame Relay networks for higher speed access lines as well as lower cost switches. Frame relaying offers this potential due primarily to the lower protocol overhead required—a recognition that current Level 3 and 4 network protocols offer effective error detection and recovery without particular regard to the underlying subnetwork architecture. Although data is passed from entry point to exit point in the network over a *Permanent Virtual Circuit* (PVC), the network does not assure error-free transmission, and thus has no requirement to retain a copy of each frame for possible retransmission. Frame switches do, however, assure delivery of frames in correct order. Frame relay promises both connection oriented and connectionless service, but only connection oriented services are being offered initially, as the specification on connectionless services is incomplete. Initial implementations allow PVCs to be administratively defined, typically between bridges or routers linking LANs across the Frame Relay network.

Addressing

A default frame format is shown in Figure 2. The *Data Link Connection Identifier* (DLCI) is used to identify the logical connection multiplexed within the physical channel, with which a frame is associated. The basic Frame Relay specification defines DLCIs with *local significance,* i.e., the same DLCIs may be simultaneously in use by different switches in a network. One of the extensions discussed later defines *global addressing,* where DLCIs are unique for each end point in the network. The minimum, and default, DLCI is two octets, ten bits of which constitute address, and six bits which carry *Extended Address* (2 bits), *Congestion* (2 bits), *Discard Eligibility* (1 bit), (a congestion option) and *Command/Response* signalling (1 bit) information. Forward Congestion and Backward Congestion information may be used with destination and source controlled transmitter rate adjustment protocols, such as well behaved TCP/IP versions. Discard Eligibility is an attempt to indicate priority, i.e., a packet which may be dropped in a congestion situation.

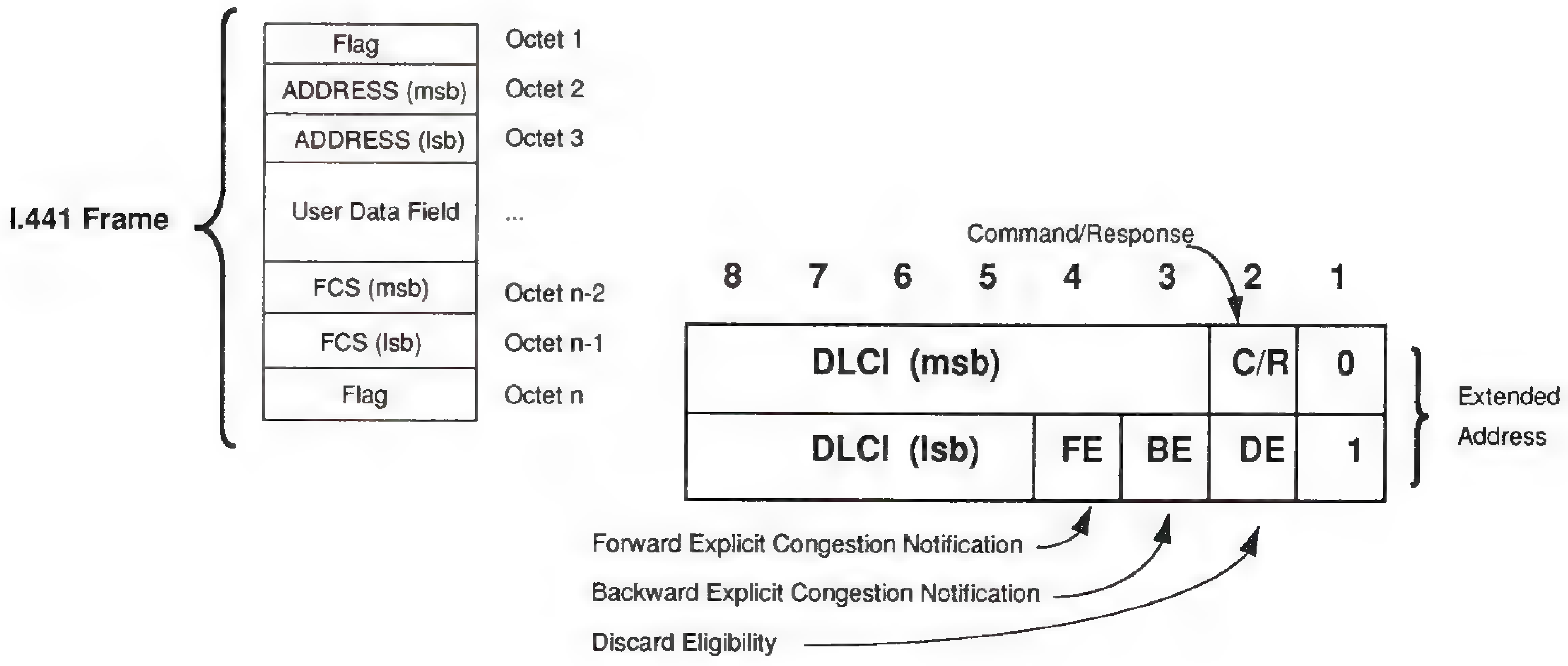


Figure 2: Frame format and addressing conventions

continued on next page

Frame Relay *(continued)*

This protocol support notwithstanding, initial implementations do not typically use the congestion bits for data virtual circuits. Both implementors and the ANSI T1S1 committee are working to better understand whether the existing specification will actually work, and how/when intermediate switches should designate links as "congested," as well as the proper actions.

The United States recommendation to CCITT is to support maximum frame sizes of up to 1600 bytes to prevent excess segmentation and reassembly by user equipment [3]. However, the actual maximum frame sizes supported by initial switch implementations varies from vendor to vendor, which may require routers to perform fragmentation of Ethernet or Token Ring LAN packets prior to transmission over the Frame Relay interface. This can have adverse performance impacts if full size Ethernet or Token Ring packets (say, NFS packets) are bridged or routed across the Frame Relay network. This is also the case, of course, for LAN interconnection across X.25 networks.

LAN interconnect issues

Using a Frame Relay network for multiprotocol LAN interconnection offers benefits and drawbacks. On the pro side, bursty LAN traffic from multiple sites may share a common access WAN, allowing more efficient use of expensive bandwidth. The combination of voice and Frame Relay for data by some switch vendors may be suitable for those organizations with a roughly equal volume of both, permitting the creation of a single WAN with integrated voice/data services. On the con side, Frame Relay is an incomplete standard which lacks much of the protocol support required to interconnect multiprotocol LANs with the performance and flexibility that current high performance bridges and routers can offer over dedicated, point to point communication links. The shortfalls are primarily in five areas:

- Level 3 PDU/Frame Relay address resolution
- Network Management functions
- Interface Management functions
- Level 3 PDU encapsulation issues
- Interior network routing

Interior network routing is a function of the specific subnetwork implementation (whatever the vendor uses for switch-switch management and routing protocols) and not visible to the user device. *Level 3 PDU encapsulation* for multiple protocols needs to be codified, but was not addressed in the joint specification, which focused on the specific interface in the Frame Relay access layer. Support for *address resolution, Network and Interface Management,* and *LAN interconnection functions* was otherwise codified in an enhanced access layer, termed the *Local Management Interface* (LMI) which is described below.

Local Management Interface

As seen in Figure 3, Frame Relay services operate at Level 2, the Data Link Level, of the OSI Reference Model. As defined by CCITT, the network does not provide any services between the DCE interface and the user device (e.g., Frame Relay switch and router). The cisco/DEC/NTI/Stratacom specification proposes a set of required LMI functions, with options, to convey information between user equipment (typically routers or bridges) and the Frame Relay network.

<table>
<tr><td colspan="2">7 - Application</td><td colspan="8"></td></tr>
<tr><td colspan="2">6 - Presentation</td><td colspan="8"></td></tr>
<tr><td colspan="2">5 - Session</td><td colspan="8"></td></tr>
<tr><td colspan="2">4 - Transport</td><td colspan="8"></td></tr>
<tr><td colspan="2">3 - Network</td><td colspan="8"></td></tr>
<tr><td rowspan="2">2 - Data Link</td><td>LLC</td><td colspan="5">802.2 Logical Link Control</td><td colspan="2">LAPD Protocol</td><td>PPP</td></tr>
<tr><td>MAC</td><td>802.1 Bridge</td><td>802.3 MAC</td><td>802.5 MAC</td><td>802.6 MAC</td><td>FDDI MAC</td><td>ATM</td><td>Frame Relay</td><td>HDLC</td></tr>
<tr><td colspan="2">1 - Physical</td><td colspan="2">802.3 PHY</td><td>802.5 PHY</td><td>802.6 PHY DS3, SONET</td><td>FDDI PHY</td><td>SONET</td><td>Synch Serial</td><td>Synch Serial</td></tr>
</table>

Figure 3: Frame Relay and the OSI Reference Model

Common extensions

As mentioned earlier, PVCs between routers/bridges are administratively added and deleted by network operators. A set of Enquiry/Status messages has been defined to allow either the network or user device to query the status of PVCs, and thus be notified upon the deletion, addition, or change in status of PVCs. Other status messages support a *keep alive function,* typically used by routers to verify the integrity of the physical link between network and user. In this instance, a sequence number is exchanged and incremented between both ends of the physical link. Thus, although no data traffic might be present, the router can determine the logical state of the network link, as well as the PVCs supporting communications across the Frame Relay network.

LMI optional extensions

Several additional functions were defined to optimize the use of routers and bridges on the Frame Relay network. These functions are options in the proposed extensions, and were designed explicitly to facilitate LAN interconnection.

The first option is *global addressing,* where each Frame Relay end point (typically a router or bridge) has a unique DLCI address. Thus, packets to be delivered to a specific end point will have the DLCI address field for that end point. When delivered at the proper destination, the delivering Frame Relay switch alters the DLCI field to indicate the *source* of the packet immediately before actual delivery to the end point. Thus, the single address field can function as a source and destination field. Typically this is achieved by using the source DLCI as an index into a table of PVCs at the entry Frame Relay switch. The PVC is actually carried through the network with the packet in the DLCI field, and is re-mapped at the exit switch into the destination DLCI.

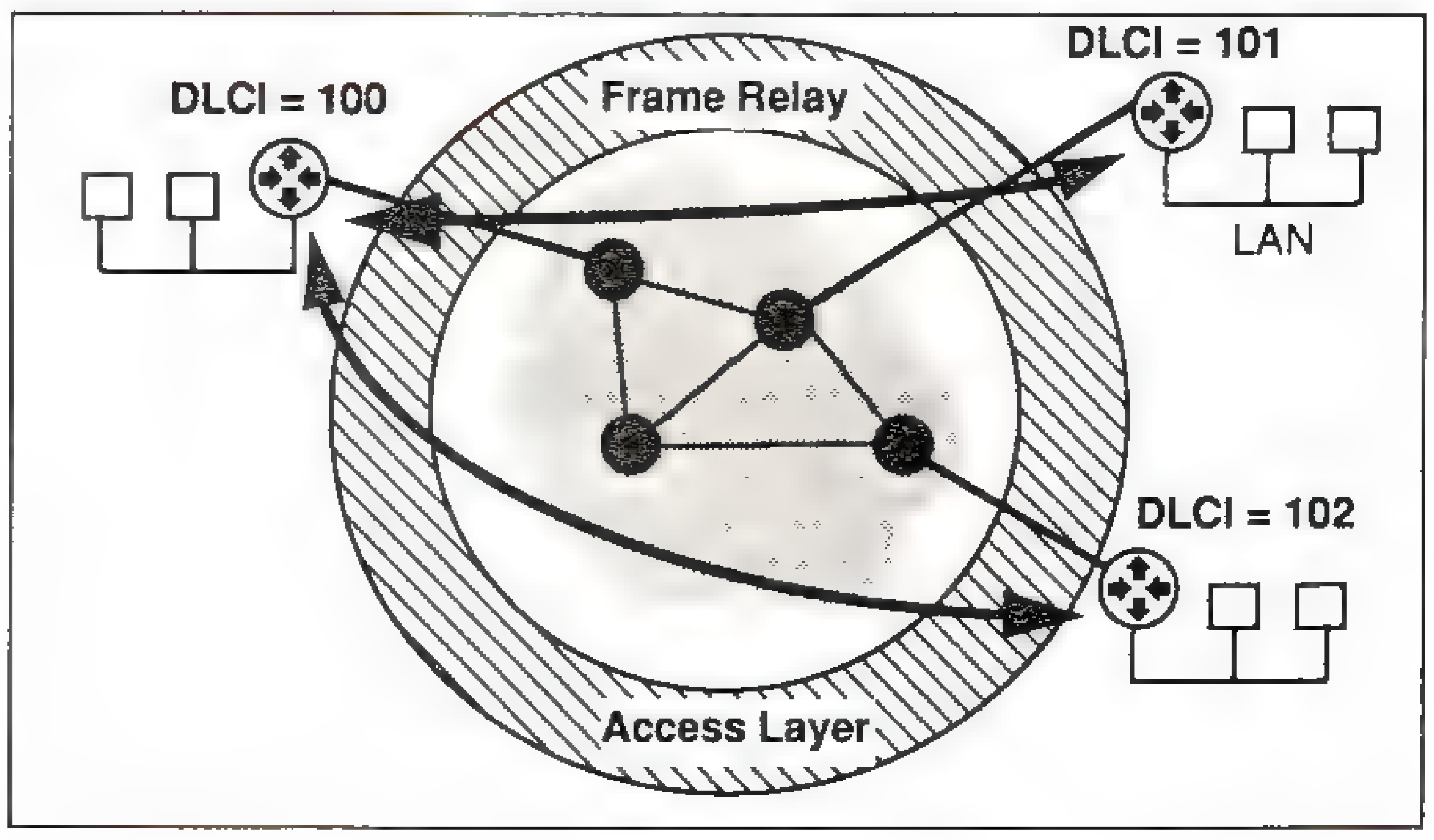


Figure 4: Global Addressing—Each end point has unique DLCI

continued on next page

Frame Relay *(continued)*

Multicast

A service was defined to allow *multicast groups* to be defined within the Frame Relay network. A multicast packet is any packet sent to a reserved DLCI; the network will replicate the packet and send a copy to all end points in a designated set. Provisions for up to four multicast groups were specified, although in practice only one is used. As with all Frame Relay packets, delivery is not guaranteed. A set of LMI messages was defined to support the configuration and monitoring of the multicast function:

- Notification of the addition, deletion and presence of a multicast group.
- Notification of the availability or unavailability of a configured multicast group. Multicast support is administratively controlled by the Frame Relay network operator, and support may be instated or removed at any time.
- Notification via an LMI message of the *source DLCI* of a multicast message. The normal practice of the Frame Relay switch of altering the DLCI field to indicate the source DLCI is not followed for multicast packets; the destination DLCI is left unaltered to indicate the multicast nature of the packet.

Address resolution

A primary use for the multicast service over the Frame Relay network is address resolution between the Level 2 (LAPD DLCI) and Level 3 (TCP/IP, DECnet, etc.) addresses. The service is analogous to the *Address Resolution Protocol* (ARP) used for Ethernet MAC to TCP/IP mapping. The presence of dynamic address resolution greatly facilitates addition or deletion of routers to a Frame Relay network, as the internal tables mapping the LAPD addresses to (for instance) IP addresses can be automatically updated with changes to network connectivity. With X.25 networks this is not possible as there is no standard multicast or broadcast service, and the maps between X.121 and Level 3 protocols must be manually updated with each addition or deletion of routers or bridges to the X.25 network.

Although address resolution is a Level 3 function and beyond the scope of the joint specification, is bears mention that there is active discussion of this topic, as many emerging network architectures require such a service. The *Switched Multimegabit Data Service* (SMDS) [7] proposed by Bellcore and the U.S. RBOCs, for example, must resolve the same problem to provide effective LAN interconnection for multiple protocol suites. One proposed solution envisions the use of well known, distributed servers to provide address resolution, and/or an extended *Point-To-Point Protocol* (PPP) [4, 5]. Lacking any standard or proposed standard, several proprietary address resolution schemes are in use to provide multiprotocol service over Frame Relay networks.

Asynchronous LMI services

Additional services were defined in the optional extensions to provide timely notification to user devices of changes to network status. The common extensions define a *Heartbeat Process,* which would typically be a software process in the router periodically polling the Frame Relay switch with *Enquiry/Status* messages. However, asynchronous notification provides the router with information faster, resulting in more timely rerouting, network problem resolution, and potentially lessening the amount of polling traffic between the router and switch as the heartbeat interval can be increased.

The following LMI services were defined in the options:

- The network will notify the user device in the case of a change in DLCI status (e.g., addition/deletion of DLCIs or PVCs).
- The network will notify the user device of the deletion of a PVC or multicast group (DLCI). In this case, the user device would regard this action as a loss of connectivity across the Frame Relay network and may make alternate routing decisions.

Bandwidth information and flow control

An enhanced PVC Status message was defined to include a field with a 24 bit value indicating the available bandwidth for a specific PVC. This information can be used by the user device, for example, to make routing decisions based on available throughput. The network, on the other hand, may allocate or deallocate "available" bandwidth based on service (e.g., time of day, charging structure, etc.) or operational considerations (loss of trunk links, bandwidth, etc.).

A further LMI service was defined to support flow control between user devices and the network. Each PVC has a configured buffer available at the switch *(nN10)*, with two thresholds, *nN11* and *nN12*. If the first threshold is exceeded, a PVC status message is sent from the switch to user device with the "R" bit set (see Figure 5). A further message is sent when the second threshold is reached with the "R" bit cleared to indicate that the buffer has emptied enough to accept additional data. Table 1 shows the applicable LMI parameters, and Figure 5 illustrates an example, complete, UPDATE STATUS message from the switch to user device. In this example, two information elements are included in the message; the first is a multicast group status message, the second is a PVC status message. Additional detail on this example can be found in the joint specification [1].

8	7	6	5	4	3	2	1	
0	1	1	1	1	1	1	0	
1	1	1	1	1	1	0	0	Management DLCI (Level 2)
1	1	1	1	0	0	0	1	
0	0	0	0	0	0	1	1	
0	0	0	0	1	0	0	1	
0	0	0	0	0	0	0	0	
0	1	1	1	1	1	0	1	
0	0	0	0	0	0	0	1	Report Type (Full Status Msg)
1								
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1	1	Keep Alive Sequence
2								
Current Sequence Number								
Last Sequence Number Received								
0	0	0	0	0	1	0	1	Multicast Status
6								
Multicast Group ID								
Multicast DLCI (msb)								
Multicast DLCI (lsb)								
Source DLCI (msb)								
Source DLCI (lsb)								
0	0	0	0	N	D	A	0	
0	0	0	0	0	1	1	1	PVC Status
6								
PVC DLCI (msb)								
PVC DLCI (lsb)								
0	0	0	0	N	D	A	R	
PVC Bandwidth (msb) [3 octets]								
FCS (msb)								
FCS (lsb)								

Figure 5: LMI Options—Sample STATUS message with multiple fields

continued on next page

Frame Relay *(continued)*

Additional LMI Parameters					
Name	*Range*	*Increment*	*Default*	*Units*	*Who*
nN10	10 - 65535	1	4000	bytes	Network
nN11	1 - 100	1	75	%	Network
nN12	1 - 100	1	25	%	Network
nP10	1	n/a	n/a	groups	Network

Table 1: LMI Parameters—Optional Extensions

Summary

Initial implementations of this interface are being tested and operated at several customer sites, and was generally available in December 1990. The joint specification has been proposed to the ANSI T1S1 committee, which is considering the specific extensions and service definitions. Several working groups within the Internet Engineering Task Force (IETF) are considering consolidating the previously separate subjects of IP over Large Public Data Networks, IP over Frame Relay, and SMDS Networking into a single group considering the problems of LAN interconnection over arbitrary large public and private network architectures. Work remains to evaluate the effectiveness of the address resolution and encapsulation techniques being first used for interconnection over Frame Relay networks, and to define standards to ensure multivendor compatibility and interoperability. The promising acceptance of the joint specification holds out the possibility that the introduction of LAN interconnection services, public and private, over Frame Relay networks can be greatly speeded up, as well as assuring potential users that, in the main, vendors are working together to reduce confusion and provide effective, efficient, and standards-based solutions.

Getting more information

cisco Systems, Inc.
Brent Bilger, 415-326-1941
1525 O'Brien Drive
Menlo Park, CA 94025

Northern Telecom, Inc.
Scott Schauer, 919-992-1390
4001 East Chapel Hill - Nelson Hwy
Research Triangle Park, NC 27709

Digital Equipment Corporation
William Mitchell, 508-486-7884
550 King Street LKG 2-1/N2
Littleton, MA 01460

Stratacom, Inc.
Brian Button, 408-370-2333
3175 Winchester Blvd.
Campbell, CA 95008

References

[1] cisco/DEC/NTI/Stratacom, "Frame Relay Specification with Extensions," September 1990. (Availble from above contacts).

[2] ANSI T1S1, "DSS1 Core Aspects of Frame Relay," March 1990.

[3] ANSI T1S1, Draft on Frame Size.

[4] Perkins, D., "Point-to-Point Protocol for the Transmission of multi-protocol datagrams over Point-to-Point links," RFC 1171.

[5] Perkins, D., Hobby R., "Point-to-Point Protocol (PPP) initial configuration options," RFC 1172.

[6] Vair, D., "Components of OSI: X.25—the Network, Data Link, and Physical Layers of the OSI Reference Model," *ConneXions,* Volume 4., No. 12, December 1990.

[7] Hughes, L., & Starliper, S., "Switched Multimegabit Data Service (SMDS)," *ConneXions,* Volume 4, No. 10, October 1990.

EDWARD KOZEL received his B.S.E.E. (1978) from U.C. Davis. For the past two years he has worked for cisco Systems, Inc. as business development manager, establishing and managing joint relationships and special projects. Before joining cisco, he worked for six years with SRI International's Information Sciences Technology Center on various projects concerned with the design and application of internetwork technologies, including Packet Radio, gateways, and ARPANET.

The Intermail Service and the Commercial Mail Relay Project

by
Ann Westine, Annette DeSchon, Jon Postel, Craig E. Ward,
University of Southern California,
Information Sciences Institute

Introduction

The evolution of large electronic mail systems testifies to the increasing importance of electronic mail as a means of communication and coordination throughout the scientific research community. This article is a summary of an experiment in protocol interoperation between mail systems of different design. USC/Information Sciences Institute (ISI) began work on this experiment in 1981 and over the years has provided an evolving demonstration service for users to exchange mail between the Internet and a few commercial mail systems.

The Internet

The *Internet* is an interconnected system of networks using the *SMTP* mail protocol, which includes the MILNET, NSFNET, and about 2000 other networks; mail relays allow the exchange of mail with BITNET, CSNET, and the UUCP networks as well. To the users, this Internet looks like one large mail system with at least 100,000 computers and at least 400,000 users.

As commercial mail systems came into popular use, it became clear that a mail link between the Internet and the commercial mail systems was necessary. The Intermail service allows these groups to communicate with Internet users by purchasing electronic mail services from commercial companies.

The experiment included the Internet mail system, the US Sprint *Telemail* system, the *MCI Mail* system, and the *Dialcom* systems. Each of these systems was originally designed to operate autonomously, with no convenient mechanism to allow users of one system to send electronic mail to users on another system.

Recently other organizations have begun to provide similar services, demonstrating the ongoing need for interoperation of the Internet and the commercial mail systems. We believe that ISI's pioneering work in this area has promoted this expansion of service.

Intermail service

Intermail is an experimental mail forwarding system for sending electronic mail across mail system boundaries. Users on each system are able to use their usual mail programs to prepare, send, and receive messages. No modifications to any of the mail programs on any of the systems are required.

The earliest version of Intermail was developed in 1981, by Jon Postel, Danny Cohen, Lee Richardson, and Joel Goldberg. It ran on the *TOPS-20* operating system and was used to forward VLSI chip specifications for the *MOSIS* project between the ARPANET and the Telemail system. Later, in 1983, Annette DeSchon converted Intermail into a more general-purpose mail-forwarding system, supporting forwarding between the Internet mail system and three commercial mail systems: Telemail, MCI Mail, and Dialcom.

Addressing

The addressing appears at the beginning of the text of each message. It can be used to include various Internet mail header fields in addition to the standard “To:” and “Cc:” address fields. This format also allows the use of special address formats, such as U.S. postal addresses and Telex addresses, which are supported by the MCI Mail system. The Intermail system performed partially automated error handling.

continued on next page

Intermail and Commercial Mail Relay *(continued)*

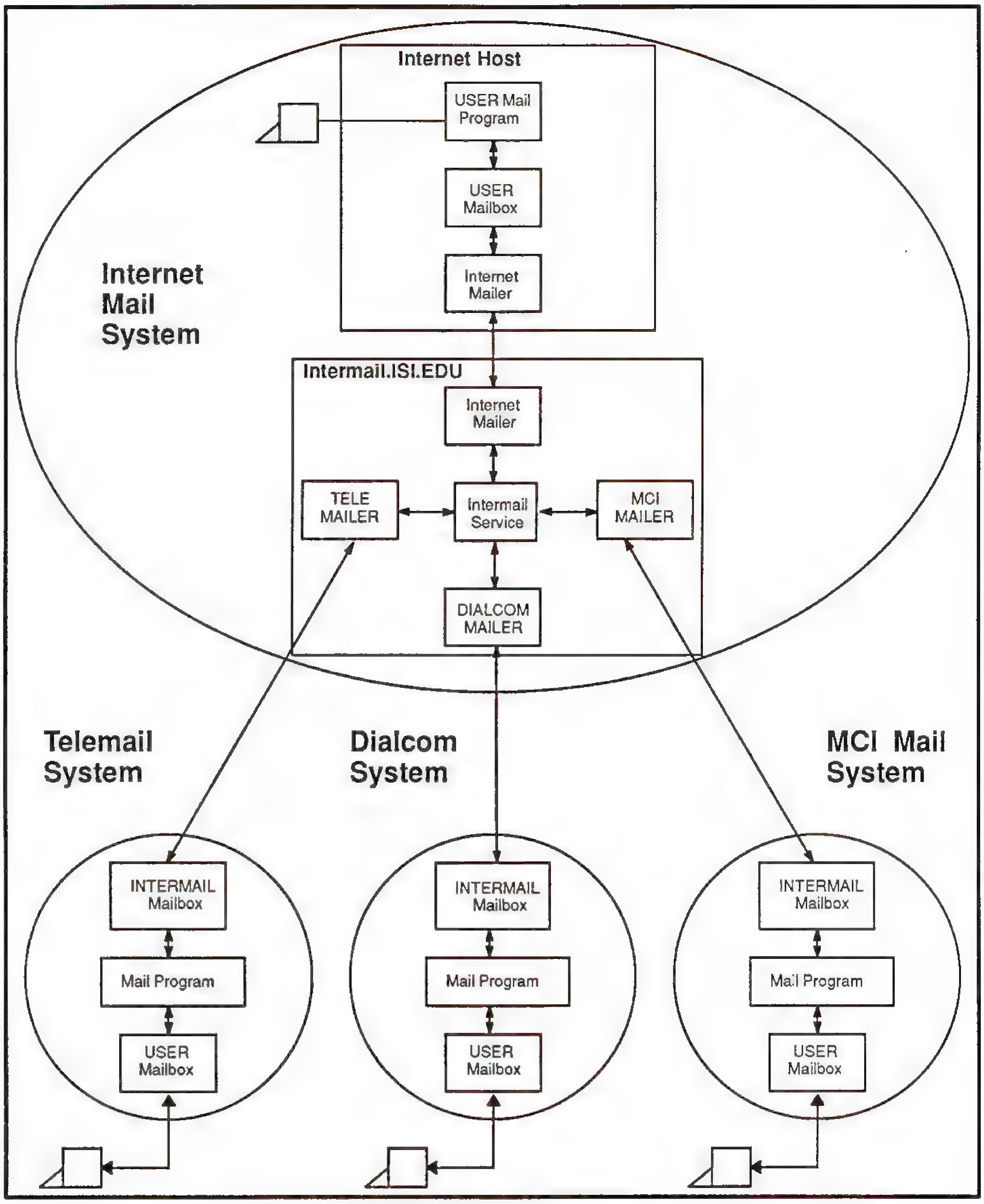


Figure 1: System Architecture

Commercial Mail Relay

In 1988, the *Commercial Mail Relay* (CMR) project developed new software to run on a dedicated UNIX system, replacing the TOPS-20-based Intermail system. The Commercial Mail Relay is being developed by Craig E. Ward. Ann Westine served as the Postmaster for both Intermail and the CMR until March 1989. Currently, the ISI Action Office serves as Postmaster.

The Intermail service now supports relay-style addressing in the Internet to commercial system direction. One advantage of relay-style addressing is that users from different commercial systems can appear on Internet mailing lists. Another advantage is that the "reply" features of most Internet user applications can be used to respond to mail that originated on a commercial system. Unfortunately, the commercial systems have no mechanism that allows users to enter Internet addresses directly into the message header. The service now supports automated error handling, which enables it to provide faster turnaround on messages containing addressing errors, and requires much less intervention from a human postmaster.

System description

The *Multi-channel Memo Distribution Facility* (MMDF) is used as the system mail software. Its notion of separating the mail queue into separate *channels* makes it easy to dedicate a channel/ queue combination (or "mailer") to each commercial system. Internet mail comes in over the standard SMTP port, and the system parses the destination address, queuing the message in the proper outgoing mailer.

The system uses a mailbox on each commercial system. Commercial users send mail to this mailbox with the address included in the message text. Each mailer, in addition to sending outgoing mail into the commercial system, reads all messages in the mailbox and places them in a queue. (See Figure 1). The processing of this queue is performed by the Intermail service. It parses each message header to find the sender and subject, then it searches for and processes the destination addresses, and routes to the outgoing mailer.

The system employs a simple accounting mechanism: a shell script counts the number of times a string marker occurs in the MMDF logs. At the end of the month, another script uses an *awk* program to total the number of messages sent and received with each commercial system.

Traffic data

The following traffic data indicates the total number of messages sent, and read for all systems during the months of March, April, and May 1990.

Month	*Messages per Day*	*Total*
March	129	3988
April	134	4022
May	118	3670

Commercial systems served

The Intermail service connects the internet mail system with the commercial systems of *Sprintmail* (formerly Telemail), MCI Mail, and Dialcom (including *Compmail, CGNET,* and *USDA-Mail*). Specific examples of the users of the Intermail services are as follows:

- Scientists in Oceanography, Astronomy, Geology, and Agriculture use Intermail to communicate with colleagues. Many of these scientists have accounts on "Sciencenet," which is actually part of a Telemail system administered by *Omnet.*
- The IEEE Computer Society's publication editors use the Dialcom system "Compmail" to manage the papers being prepared for their numerous publications. Many of the authors are in university departments with access to the Internet. Intermail support a significant exchange of large messages containing manuscripts.
- NASA uses Telemail systems for its own work and has extensive exchanges through its own relay service, as well as Intermail, for communicating with university scientists on the Internet.

Acceptable Use Policy

The Internet is composed of many networks sponsored by many organizations. However, most of the Internet's long-haul networks are provided by U.S. government agencies. Each of these agencies limits the use of the facilities it provides in some way. In general, the statement by an agency about how its facilities may be used is called an "Acceptable Use Policy." The Intermail service at ISI is a resource provided by the *Defense Advanced Research Projects Agency* (DARPA) for computing and communication. Use of this resource must be limited to DARPA-sponsored work or other approved government business or must otherwise meet the Acceptable Use Policy of DARPA.

continued on next page

Intermail and Commercial Mail Relay *(continued)*

However, DARPA, as a member of the *Federal Networking Council* (FNC), has said that: "The member agencies of the FNC agree to carry all traffic that meets the Acceptable Use Policy of the originating member agency." In the least restrictive case, all bona fide researchers and scholars, public and private, from the US and foreign countries (unless denied access by national policy) are allowed access.

BITNET and UUCP (and other) networks are interconnected to the Internet via mail relays. It is the responsibility of the managers of these mail relays to ensure that the e-mail messages ("traffic") that enter the Internet via their mail relays meet the Acceptable Use Policy of the member agency providing the Internet access. In addition, we cannot allow the Intermail service to be used simply as a bridge between two commercial systems, even though the system has this technical capability. At least one end of the communication must be related to FNC acceptable use.

Other mail relays

Recently, other groups have begun to offer mail relaying between the Internet and some commercial mail systems. This demonstrates the need for interoperation between the Internet and the commercial services. The success of the Intermail service has encouraged the development of these additional relays. Potential users are encouraged to contact any of the following services.

- *The Intermail gateway to Sprintmail, MCI Mail and Dialcom:*
  `Intermail-Request@INTERMAIL.ISI.EDU` 1-213-822-1511
- *The Merit gateway to Sprintmail and IEEE Compmail:*
  `Customer.Service@SPRINT.COM` 1-800-336-0437
- *The CNRI gateway to MCI Mail:*
  `0002671163@MCIMAIL.COM` 1-800-444-6245
- *The Ohio State University gateway to Compuserve:*
  `Postmaster@CSI.COMPUSERVE.COM` 1-800-848-8990
- *NASA-Ames gateway to Telemail:*
  `admin/arc@nasamail.nasa.gov` 1-415-694-4180

The Internet and Connected Networks

The Internet is a network of networks interconnected by *gateways* or *routers*. The common element is the TCP/IP protocol suite. It now includes approximately 2000 networks and 100,000 host computers. The Internet is made up of local area networks in research institutes and university campuses, regional networks, and long-haul networks. These resources are supported by the using organizations and by several US government agencies (including DARPA, NSF, NASA, DoE, and NIH). The long-haul networks in the Internet are the *MILNET*, the *NSFNET* Backbone, the *NASA Science Internet* (NSI), and the DoE *Energy Science Network* (ESNET).

Other systems using TCP/IP or other protocols may be networks of networks or "internets" with a lower case "i." The capital "I" Internet is the one described above. There are other networks with (semi-) compatible electronic mail systems. These include BITNET/CSNET, UUCP, ACSNET, and JANET. Users of electronic mail may not necessarily be aware of the boundaries between these systems and the Internet.

The Domain Name System

The *Domain Name System* (DNS) provides for the translation between host names and addresses. Within the Internet, this means translating from a name, such as ABC.ISI.EDU, to an IP address such as "128.9.0.123," The DNS is a set of protocols and databases. The protocols define the syntax and semantics for a query language to ask questions about information located by DNS-style names. The databases are distributed and replicated. There is no dependence on a single central server, and each part of the database is provided in at least two servers.

In addition to translating names to addresses for Internet hosts, the DNS provides for registering DNS-style names for other hosts reachable (via e-mail) through gateways or mail relays. The records for such names point to an Internet host (one with an IP address) that acts as a mail forwarder for the registered host. For example, the Australian host yarro.oz.au is registered in the DNS with a pointer to the mail relay uunet.uu.net.This gives electronic mail users a uniform mail addressing syntax and avoids making them aware of the underlying network boundaries.

References

[1] DeSchon, A., "MCI Mail/ARPA Mail Forwarding," ISI/RR-84-141, August 1984.

[2] DeSchon, A., "Intermail, An Experimental Mail Forwarding System," ISI/RR-85-158, September 1985.

[3] Westine, A., A. DeSchon, J. Postel, & C. Ward, "Intermail and Commercial Mail Relay Services," ISI/RR-90-254, March 1990.

[4] Westine, A., A. DeSchon, J. Postel, & C. Ward, "Intermail and Commercial Mail Relay Services," RFC 1168, July 1990.

[5] Westine, A., A. DeSchon, J. Postel, & C. Ward, "Intermail and Commercial Mail Relay Services," SIGUCCS Conference, October 1990.

ANN WESTINE is a 10 year veteran research staff member working on the Internet Concepts Project at University of Southern California/Information Sciences Institute (USC/ISI). She is currently involved in communications and network administration. Ann administers the US Domain, is the information contact for Los Nettos, maintains the IAB task force mailing lists. Through the years she has consoled many users on the practice of sending e-mail, to and from the Internet and commercial systems. Ann has prepared the Internet Monthly report since its beginning in 1984, and has written several articles for the network community. She can be reached as Westine@ISI.EDU.

ANNETTE DeSCHON is a member of the research staff at USC/ISI. Her interests include network protocols, internetworking and electronic mail. She received her B.A. from the University of California, Los Angeles, in 1975. She can be reached as DeSchon@ISI.EDU.

JON POSTEL is Director of the Communications Division of USC/ISI. Jon has been involved in the development of computer communication protocols and applications from the early days of the ARPANET. His current interests include multimachine internetwork applications, multimedia conferencing and electronic mail, very large networks, and very high speed communications. Jon received a BS and MS in Engineering and a PhD in Computer Science from the University of California, Los Angeles. He can be reached as Postel@ISI.EDU.

CRAIG E. WARD received his B.A (1979) from the University of California, Irvine, an A.S. (1983) from Santa Ana College and has begun work towards a M.S. at the University of Southern California. He has worked for USC/ISI since 1982 and has developed or maintained a variety of applications software for TOPS-20 and UNIX. His primary community involvement is with the local chapter of the National Space Society. He can be reached as Ward@ISI.EDU.

Book Review

X Window System Toolkit: The Complete Programmer's Guide and Specification, by Paul J. Asente and Ralph R. Swick, with Joel McCormack. Published by Digital Press, ISBN 1-55558-051-3, 1990.

Background

Developed at MIT in the late '80s, with major contributions from industry and academia, the *X Window System* has gained widespread acceptance in the workstation industry. It has been adopted by almost every major vendor as the primary display system technology. It is—like Fortran and TCP/IP—destined to be with us for years to come.

With the growing importance of X, several books and a flood of magazine articles have been published recently to bring programmers "up to speed" in using this new technology. From the perspective of programmers developing applications with graphical user interfaces, the standard *Xlib* interface is necessary, but not sufficient. Alone, *Xlib* is too cumbersome and awkward to be practical. The variety of toolkits and interface builders, which raise the level of abstraction used by the programmer, attest to the need for better application development tools.

But, underlying nearly all of these higher-level tools are the *X Intrinsics:* the software layer that provides the "substrate" for organizing the principal user interface component, the "widget." For toolkits organized around widgets, including OSF *Motif* and the latest versions of the Sun/AT&T *Open Look* toolkits, the *X Intrinsics* form a common basis. Thus, the study of the *X Intrinsics* is important to anyone wishing to know how all these toolkits work.

Organization

Enter the *X Window System Toolkit* by Asente, Swick, and McCormack. This book, written by three of the principal architects of the *X Intrinsics,* documents their work and their decisions, and so is the *definitive* book on the underpinnings of all toolkits based on the X Window System Intrinsics. While much of the book is routine documentation of procedural interfaces, there are a number of interesting sections scattered about that describe the history and background of the group and the rationale for their decisions.

It is a long book, nearly 1000 pages, and is divided into two major parts: a programmer's guide and specifications. The programmer's guide is intended to be a complete introduction to using Intrinsics-based toolkits. The specifications part is less introductory and includes additional information necessary for someone to implement a version of the Intrinsics that conforms to the *X Consortium* standard.

The programmer's guide, however, is anything but tutorial in nature. It is organized in a similar fashion to the documentation that has been shipping with the MIT X distribution, in which sections on using widgets are intermingled with sections on writing widgets.

Furthermore, the procedures are not presented in any useful order. For example, rarely-used, low-level procedures for creating application contexts, opening the display connection, initializing the display, and creating the top-level "shell" are presented before the single general-purpose procedure that most application writers will use to accomplish the same task. As a related example, many of the essential utility procedures are not documented until halfway through the book. The large number of forward references make flipping from section to section a necessary part of reading through the material.

Complete X Toolkit Reference

Despite these shortcomings for the novice toolkit programmer, the book is chock full of useful information and advice. It is the kind of reference that an experienced toolkit programmer will consult to find answers to the subtle problems that arise when writing a sophisticated or novel application. X application developers just wondering "How does it work?" will also find this book a useful and interesting resource.

For someone extending a toolkit, by subclassing existing widget classes, this book provides the kind of detailed information that explains not only how to do it, but why it must be done that way. The last chapter of the programmer's guide provides complete examples that illustrate the details involved in writing pop-up widgets, developing sophisticated composite and constraint widgets, and managing gadgets.

This book is a must for programmers who want—or need—to know what is actually going on inside the toolkits.

—John T. Korb, Purdue Computer Science Department

More on X

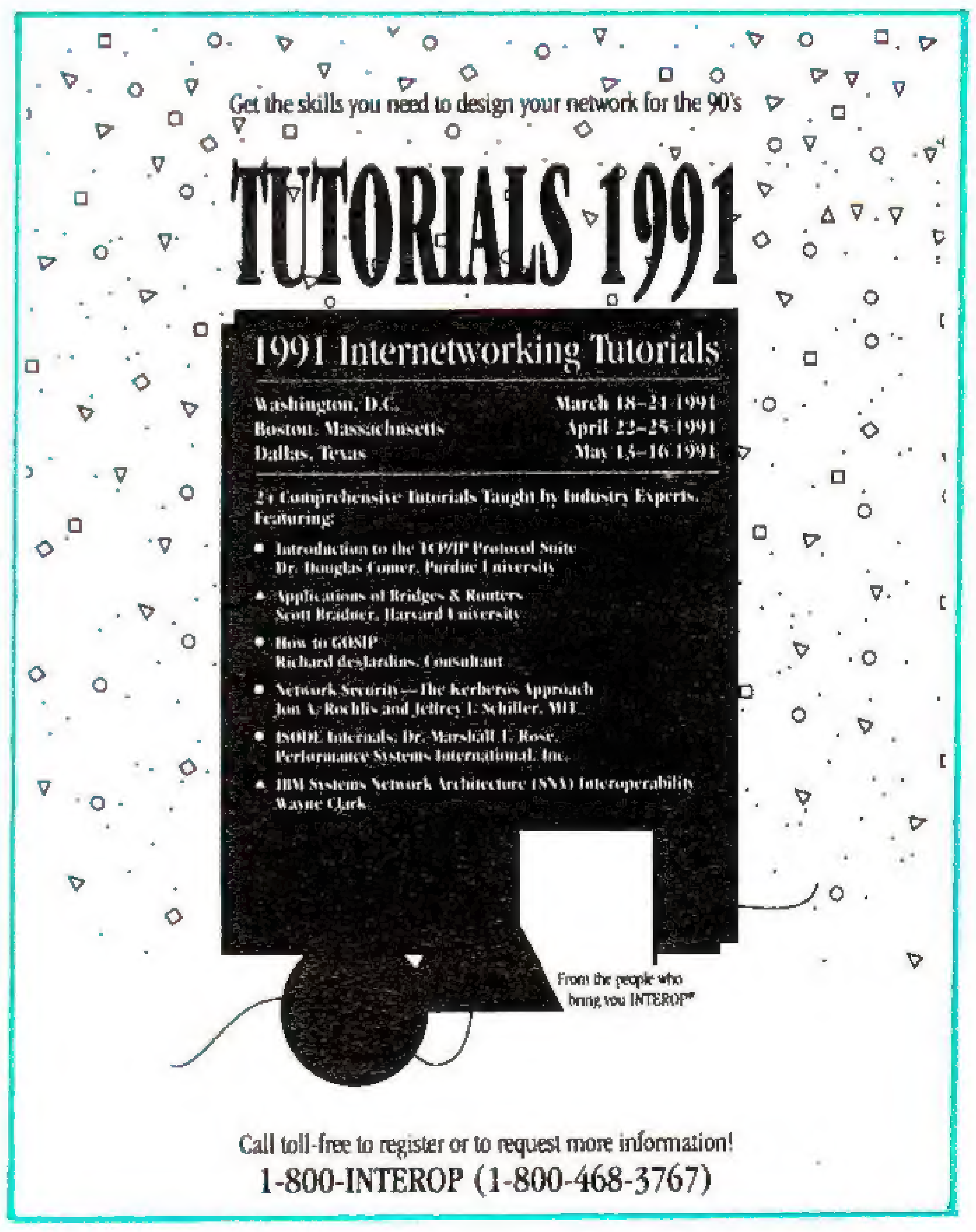
[Ed.: Two Other X Window System books were reviewed in the October 1989 issue of ConneXions. Articles on X have also appeared in this journal: "X Windows: More than Just a Pretty Face," by Bill Jolitz, ConneXions, Volume 4, No. 5, May 1990, and "A Programmer's Overview of X," by John T. Korb and Wayne Dyksen, ConneXions, Volume 4, No. 10, October 1990.]

1991 Internetworking Tutorials

The 1991 *Internetworking Tutorials Program Guide* is now available. Choose from over 20 courses in the field of networking. All these two-day classes are given by experts in the field. Call us now at 1-415-941-3399 or toll-free 1-800-INTEROP and ask for the complete tutorials guide. These courses will be given at 3 different locations in the next 3 months:

Dates and locations

March 18–21, 1991:	*Washington*
April 22–25, 1991:	*Boston*
May 13–16, 1991:	*Dallas*



Get the skills you need to design your network for the 90's

TUTORIALS 1991

1991 Internetworking Tutorials

Washington, D.C.	March 18–21 1991
Boston, Massachusetts	April 22–25 1991
Dallas, Texas	May 13–16 1991

24 Comprehensive Tutorials Taught by Industry Experts, Featuring:

- Introduction to the TCP/IP Protocol Suite
Dr. Douglas Comer, Purdue University
- Applications of Bridges & Routers
Scott Bradner, Harvard University
- How to GOSIP
Richard desJardins, Consultant
- Network Security—The Kerberos Approach
Jon A. Rochlis and Jeffrey I. Schiller, MIT
- ISODE Internals, Dr. Marshall T. Rose, Performance Systems International, Inc.
- IBM Systems Network Architecture (SNA) Interoperability
Wayne Clark

From the people who bring you INTEROP™

Call toll-free to register or to request more information!
1-800-INTEROP (1-800-468-3767)

CONNEXIONS
480 San Antonio Road
Suite 100
Mountain View, CA 94040
415-941-3399
FAX: 415-949-1779

ADDRESS CORRECTION
REQUESTED

CONNEXIONS

EDITOR and PUBLISHER Ole J. Jacobsen